Two Proteins Form a Heteromeric Bacterial Self-Recognition Complex in Which Variable Subdomains Determine Allele-Restricted Binding

RESEARCH ARTICLE

Two Proteins Form a Heteromeric Bacterial Self-Recognition Complex in Which Variable Subdomains Determine Allele-Restricted Binding

Lia Cardarelli, Christina Saak, Karine A. Gibbs

Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA

ABSTRACT **Self- versus nonself-recognition in bacteria has been described recently through genetic analyses in multiple systems; however, understanding of the biochemical properties and mechanisms of recognition-determinant proteins remains limited. Here we extend the molecular and biochemical understanding of two recognition-determinant proteins in bacteria. We have found that a heterotypic complex is formed between two bacterial self-recognition proteins, IdsD and IdsE, the genes of which have been shown to genetically encode the determinants for strain-specific identity in the opportunistic bacterial pathogen *Proteus mirabilis*. This IdsD-IdsE complex forms independently of other *P. mirabilis*-encoded self-recognition proteins. We have also shown that the binding between IdsD and IdsE is strain- and allele-specific. The specificity for interactions is encoded within a predicted membrane-spanning subdomain within each protein that contains stretches of unique amino acids in each *P. mirabilis* variant. Finally, we have demonstrated that this *in vitro* IdsD-IdsE binding interaction correlates to *in vivo* population identity, suggesting that the binding interactions between IdsD and IdsE are part of a cellular pathway that underpins self-recognition behavior in *P. mirabilis* and drives bacterial population sociality.**

IMPORTANCE **Here we demonstrate that two proteins, the genes of which were genetically shown to encode determinants of self-identity in bacteria, bind *in vitro* in an allele-restricted interaction, suggesting that molecular recognition between these two proteins is a mechanism underpinning self-recognition behaviors in *P. mirabilis*. Binding specificity in each protein is encapsulated in a variable region subdomain that is predicted to span the membrane, suggesting that the interaction occurs in the cell envelope. Furthermore, conversion of binding affinities *in vitro* correlates with conversion of self-identity *in vivo*, suggesting that this molecular recognition might help to drive population behaviors.**

Received 12 February 2015 **Accepted** 18 May 2015 **Published** 9 June 2015

Citation Cardarelli L, Saak C, Gibbs KA. 2015. Two proteins form a heteromeric bacterial self-recognition complex in which variable subdomains determine allele-restricted binding. mBio 6(3):e00251-15. doi:10.1128/mBio.00251-15.

Editor Richard Gerald Brennan, Duke University School of Medicine

Address correspondence to Karine A. Gibbs, kagibbs@mcb.harvard.edu.

The ability to distinguish self from foreign is found broadly in biology and is at the heart of many group behaviors. In eukaryotes, cell-type-specific protein complexes can mediate self-identity. For example, the self-recognition system of the social amoeba *Dictyostelium discoideum*, which is responsible for the separation of nonidentical strains during fruiting body development, minimally consists of two proteins, TgrB1 and TgrC1, which bind one another (1). TgrB1 and TgrC1 are highly conserved across different *D. discoideum* strains; however, each contains regions of increased variability between strains (2). Strains with noncognate TgrB1/TgrC1 pairs show separation during *D. discoideum* development (1). For bacteria, homotypic, single-protein interactions have been recently implicated in self-recognition. A surface protein, TraA, mediates the strain-specific transfer of outer membrane proteins in the social bacterium *Myxococcus xanthus*. Since transfer of surface proteins can modulate motility in this bacterium, this has important implications for coordinating social behaviors (3–5). The residues for specificity are found within the variable region (VR) of TraA in *M. xanthus* and are sufficient to establish strain identity (6).

Here we address the challenge of understanding the biochemical mechanisms that define self-identity by focusing on the interactions between two known bacterial self-identity proteins. Populations of the bacterium *Proteus mirabilis* merge with populations consisting of genetically identical cells (i.e., "self") to make one larger swarm and form visible boundaries with populations of nonisogenic cells (i.e., "nonself") when migrating over nutrient-rich surfaces (7). Boundary formation is hypothesized to result from self-recognition-mediated events, including cell death, and is likely regulated by multiple cellular and environmental factors (8–12). This territorial exclusion requires cell-cell contact (9), and in at least one strain, three gene clusters (termed *idr*, *tss*, and *ids*) comprise the necessary components (11, 13). The *tss* and *idr* genes, which encode a type VI secretion (T6S) system and putative cytotoxic elements, respectively, are needed for competition with other *P. mirabilis* strains and appear to evoke contact-dependent growth inhibition (13) similar to the *tss* and *pef* genes described for a second strain of *P. mirabilis*, HI4320 (8). Indeed, conserved T6S systems have been identified in numerous Gram-negative bacteria, where the T6S system has been shown to translocate cargo proteins from the donor cell cytoplasm into the periplasm or cytoplasm of target cells (reviewed in reference 14), and most T6S systems have been primarily described as a mechanism for competitive killing between bacteria upon physical contact (15–18). However, in discordance with the current paradigm for T6S-related genes, the third self-

recognition gene cluster, *ids*, does not contribute to competitions against other *P. mirabilis* strains and instead encodes proteins necessary for nonlethal interactions within a clonal population (11, 13). Studying the *ids* locus, therefore, addresses mechanisms for recognition of self (kin) cells as opposed to the inhibition of nonself cells.

Insights into the functions of the Ids proteins have arisen primarily from genetic and bioinformatic analyses. Briefly, deletion of the six *ids* genes, *idsABCDEF*, is sufficient to prevent *P. mirabilis* strain BB2000 from merging with an otherwise genetically identical parent population, indicating that the Ids system is necessary for establishment of self-identity in this strain (11). The introduction of *ids* genes (or of just *idsDEF*) from strain BB2000 and driven by the native *ids* promoter (also from BB2000) into strain HI4320 is sufficient to cause the recipient HI4320 strain to form a boundary with both parental strains, indicating the formation of a new identity (11). Likewise, in *trans* expression of the *ids* genes from strain HI4320 in strain BB2000 is sufficient to alter self-identity and induce boundary formation (11). Therefore, expression of foreign *ids* genes in either of these *P. mirabilis* strains can confer a new identity.

Both the *idsD* and *idsE* genes, which are carried adjacently, must originate from the same strain for two *P. mirabilis* swarming populations to merge and for a conversion of strain identity to occur (11). Compared across several *P. mirabilis* strains, the *idsD* and *idsE* genes each contain a region of reduced sequence conservation that is flanked by highly conserved sequences (11). Together, these data suggest that the *idsD* and *idsE* genes encode the information determining self-identity (11). Neither IdsD nor IdsE has a known function outside self-recognition-dependent boundary formation. Prior to this current research, there was no structural or localization prediction for either IdsD or IdsE, and no interaction partners have been described for either protein. Therefore, we hypothesized that IdsD and IdsE may comprise a complex whose function is to convey and/or determine self-identity within a bacterial population. Here we examine the *in vitro* protein-protein interactions between IdsD and IdsE, as well as determine the allele and strain specificity of the IdsD-IdsE binding interaction, including critical residues that contribute to binding specificity. We also provide evidence that *in vitro* binding affinities positively correlate with *in vivo* self-identity.

RESULTS

IdsD and IdsE interact with each other. Given the limited information available about either IdsD (1,034 amino acids) or IdsE (312 amino acids), we explored the predicted domains of these proteins. The N-terminal domain of IdsD, from amino acids 1 to 750, is predicted to consist of several components: a disordered region, a putative T6S-associated motif (19), a coiled-coil region, and a series of sequential alpha-helices (Fig. 1A). The IdsD C-terminal domain is predicted to contain two transmembrane domains from approximately amino acids 695 to 708 and 786 to 814 and an unstructured C-terminal tail from amino acids 815 to 1034 (Fig. 1A). Likewise, IdsE appears to consist of two discernible domains: an N-terminal domain that is predicted to contain two transmembrane domains from approximately amino acids 61 to 78 and 154 to 171 and a C-terminal domain (amino acids 172 to 312) that is predicted to contain six beta-strands and two alpha-helices (Fig. 1A). From this analysis, we proceeded with the hypothesis that IdsD and IdsE are membrane-associated proteins and therefore may be localized to similar cellular environments.

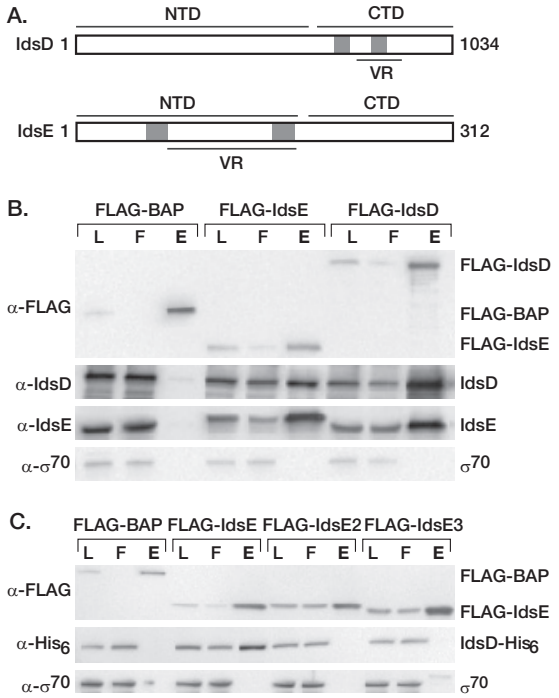

FIG 1 IdsD binds the adjacently encoded IdsE. (A) Schematic diagrams of IdsD and IdsE. Gray boxes indicate transmembrane (TM) helix predictions (IdsD amino acids 695 to 708 and 786 to 814 and IdsE amino acids 61 to 78 and 154 to 171) that were performed using TMFinder (36). Fig. S1 and Fig. S2 in the supplemental material show alignments of the full protein sequences (22–25). N-terminal domains (NTD), C-terminal domains (CTD), and variable regions (VR) are indicated by lines above or below the diagrams. The pairwise identities between variants are <70% for the IdsD VR and <40% for the IdsE VR. (B) FLAG-IdsD and FLAG-IdsE were separately immunoprecipitated from *P. mirabilis* whole-cell extract using anti-FLAG antibody resin. Samples of the experimental load (L), nonbinding fraction (F), and elution (E) were separated by SDS-PAGE and analyzed by Western blotting using anti-FLAG (top), anti-IdsD (middle), anti-IdsE (middle), and anti-σ^{70} (bottom) antibodies. The Ids-specific antibodies were verified using extracts from *P. mirabilis* swarms (see Fig. S4 in the supplemental material). *P. mirabilis* extract from the Δ*ids* strain carrying $pIds_{BB}$, which produces no FLAG-tagged proteins, was supplemented with purified FLAG-BAP and similarly analyzed. (C) Variants of FLAG-IdsE and IdsD-His_6 were separately expressed in *E. coli*, subjected to pulldown assays using anti-FLAG antibody resin, and then analyzed by SDS-PAGE and Western blotting using anti-FLAG, anti-His_6, and anti-σ^{70} antibodies (top to bottom).

We next queried whether IdsD and IdsE encoded by strain BB2000 would interact with one another. We constructed plasmids containing the entire *ids* operon under the control of the native promoter with a FLAG-tagged epitope attached in-frame to either *idsD* or *idsE* and then introduced the plasmids individually into a strain containing a chromosomal deletion of the *ids* genes (Δ*ids*). The resultant strains exhibited a wild-type boundary phenotype, indicating that the gene fusions with the FLAG epitope were functional (see Fig. S3 in the supplemental material [20]). We then performed pulldown assays with cellular extracts derived from actively migrating cells expressing either FLAG-tagged IdsD or FLAG-tagged IdsE using anti-FLAG antibody resin and analyzed the precipitates by Western blot analysis using IdsD- and IdsE-specific antibodies. IdsE was pulled down at higher levels by FLAG-IdsD than by the control protein, FLAG-BAP (Fig. 1B). Similarly, IdsD was present in the FLAG-IdsE sample, with only

trace amounts in the control (Fig. 1B). Since IdsD and IdsE co-immunoprecipitated, we conclude that they interact in *P. mirabilis* extracts.

IdsD and IdsE bind to each other independently of other Ids proteins. IdsD and IdsE binding might be direct or, alternatively, may require the other Ids proteins. To distinguish between these possibilities, we expressed both proteins individually in *Escherichia coli* strain BL21(DE3), which does not contain any *ids*-like genes. We engineered a His_6 epitope tag on the C terminus of IdsD, expressed IdsD-His_6 and FLAG-IdsE in separate *E. coli* strains, and performed anti-FLAG pulldown assays on mixed cell extracts followed by Western blot analysis. IdsD-His_6 was efficiently detected in the immunoprecipitate of FLAG-IdsE but not in that of the control protein, FLAG-BAP (Fig. 1C). Therefore, the interaction between IdsD and IdsE does not require other Ids proteins and is most likely direct.

IdsD binds specifically to the adjacently encoded IdsE variant. In several sequenced *P. mirabilis* genomes, multiple alleles of *idsE* can be identified; however, we have observed at most only one allele of *idsD* (see Table S1 in the supplemental material). For example, strain BB2000 contains two additional *idsE* alleles immediately downstream of the *idsABCDEF* gene cluster (see Fig. S5 in the supplemental material), as well as two other alleles at a distant chromosomal position, all of which have unique sequences (21). Two hypotheses for the binding interactions between IdsD and IdsE are that (i) IdsD can bind any IdsE variant found within a single genome or (ii) IdsD can bind only the IdsE protein encoded immediately adjacent on the genome. To distinguish between these hypotheses, we queried whether either of the IdsE variants encoded directly downstream of *ids* would bind IdsD-His_6 *in vitro*. Each IdsE variant (IdsE2 and IdsE3) was individually fused to FLAG, separately expressed in *E. coli*, and subjected to an anti-FLAG immunoprecipitation in the presence of IdsD-His_6. Neither downstream IdsE variant was able to effectively bind IdsD-His_6 (Fig. 1C). Taken together, we conclude there is a restricted specificity to the IdsD and IdsE binding.

IdsD-IdsE binding is strain-specific. To further examine the hypothesis that IdsD-IdsE binding specificity is restricted, we replaced *idsD* and *idsE* in the *E. coli* expression plasmids with adjacently carried alleles (see Table S1 in the supplemental material) from the independent *P. mirabilis* strains HI4320 and CW977, resulting in the production of $IdsD_{HI}$-His_6, FLAG-$IdsE_{HI}$, $IdsD_{CW}$-His_6, and FLAG-$IdsE_{CW}$. Strains BB2000, HI4320, and CW977 form boundaries against one another and recognize each other as nonself (11, 12). Anti-FLAG pulldown assays were performed on all samples. Trace levels of $IdsD_{HI}$-His_6 were detectable in pulldown assays with the BB2000-originated variant FLAG-$IdsE_{BB}$, whereas $IdsD_{HI}$-His_6 was readily detected in the FLAG-$IdsE_{HI}$ sample (Fig. 2A). Conversely, FLAG-$IdsE_{HI}$ only bound trace levels of the BB2000-originated variant $IdsD_{BB}$-His_6 (Fig. 2A). We observed similar results with $IdsD_{CW}$-His_6, which was only co-immunoprecipitated by FLAG-$IdsE_{CW}$ and not by FLAG-$IdsE_{BB}$ or FLAG-$IdsE_{HI}$ (Fig. 2B). Therefore, we conclude that the IdsD-IdsE binding interaction is primarily strain-specific.

The variable regions of IdsD and IdsE mediate binding specificity. Alignments between IdsD and IdsE variants highlight a region of high sequence variability in each protein; therefore, the strain-specific binding might be determined by these distinctive amino acid sequences. As such, we replaced the variable region in IdsE from strain BB2000 (amino acids 147 to 169) with the analogous

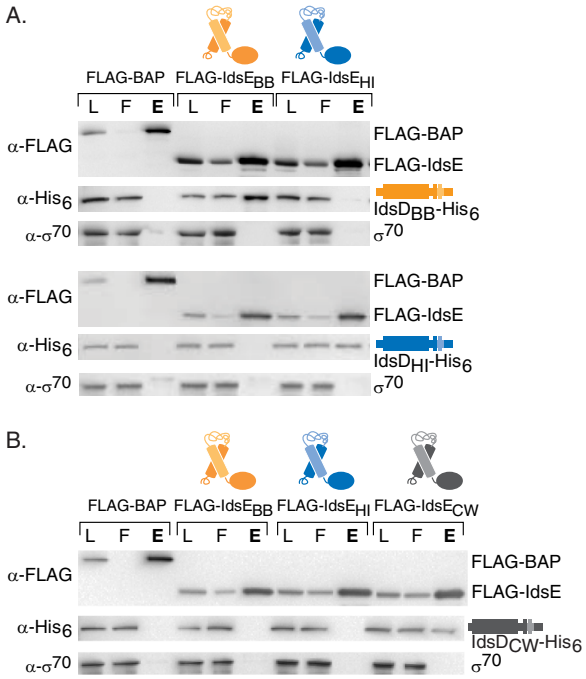

FIG 2 IdsD and IdsE exhibit strain-specific binding. Variants of FLAG-IdsE and IdsD-His_6 were subjected to pulldown assays and analyzed as described in the legend to Fig. 1. Interactions were tested between FLAG-IdsE and IdsD-His_6 originating from strain BB2000 or HI4320 (A) and FLAG-IdsE originating from BB2000, HI4320, or CW977 and IdsD-His_6 originating from strain CW977 (B). Representations of IdsD and IdsE are based on protein secondary structure predictions (Fig. 1): for IdsE, two transmembrane alpha-helices (rods) and a C-terminal domain (oval) and for IdsD, an N-terminal domain (long rectangle) and two transmembrane helices (smaller rectangles). Colors correspond to the source gene; the lighter shades indicate variable regions.

sequence from strain HI4320, resulting in FLAG-$IdsE_{BB\ to\ HI}$. We expressed this protein in *E. coli* and assayed for binding interactions by immunoprecipitation with anti-FLAG antibodies. In FLAG-$IdsE_{BB\ to\ HI}$ pulldowns, we did not detect $IdsD_{BB}$ but did detect $IdsD_{HI}$ (Fig. 3A), indicating that the variable region of IdsE is sufficient to define allele-specific binding with IdsD.

Similarly, we probed for which residues in IdsD mediate binding specificity. The 33 amino acids of IdsD unique to either the BB2000 or HI4320 variants are located across the primary sequence (see Fig. S1 in the supplemental material [22–25]). Therefore, we constructed a hybrid epitope-tagged IdsD protein, $IdsD_{large\ BB\ to\ HI}$-His_6, consisting of amino acids 1 to 442 and 866 to 1034 of $IdsD_{BB}$, where 4 amino acid polymorphisms lay, and amino acids 443 to 865 of $IdsD_{HI}$, which contained the remaining 29 polymorphisms and is part of the C-terminal domain (Fig. 4). We expressed this protein in *E. coli* and assayed for binding interactions by anti-FLAG immunoprecipitation and found that hybrid $IdsD_{large\ BB\ to\ HI}$-His_6 was pulled down robustly by both FLAG-$IdsE_{HI}$ and FLAG-$IdsE_{BB\ to\ HI}$ but not by FLAG-$IdsE_{BB}$ (see Fig. S6 in the supplemental material). This result suggested that the residues sufficient to convert IdsD to a new binding specificity were found in the C-terminal domain.

To more narrowly define the variable region of IdsD, we reasoned that the variable region we identified for IdsE mapped to a region predicted to be surrounding a membrane-spanning portion; therefore, we predicted that the complementary surface on IdsD would map to a similar position (Fig. 4). As such, we replaced

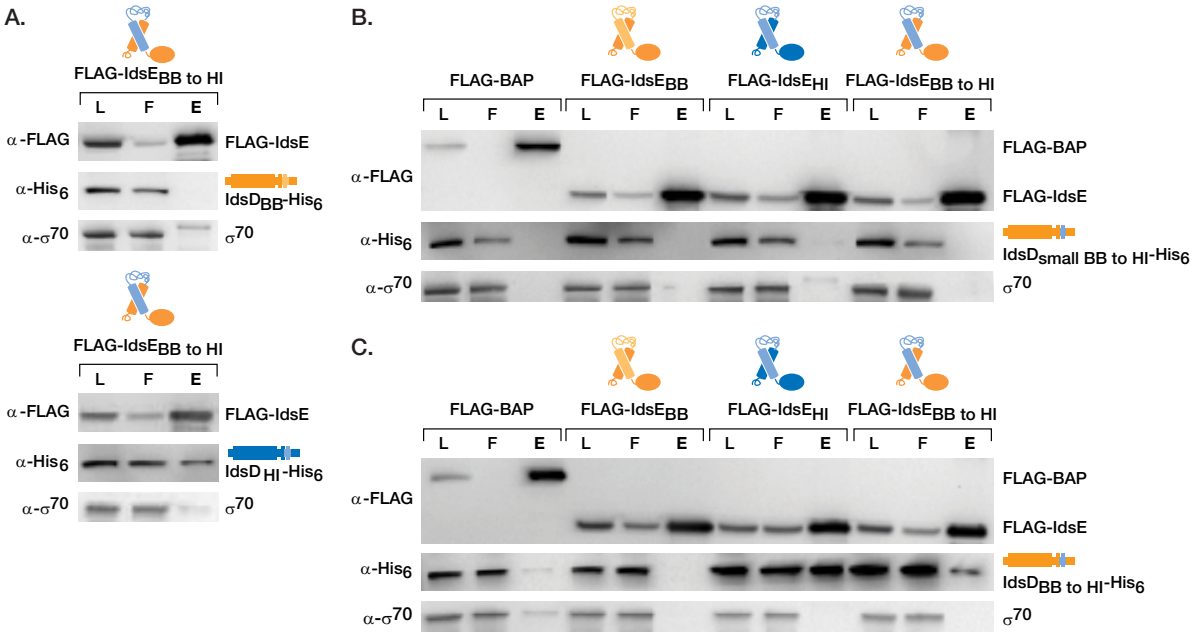

FIG 3 The variable regions of IdsD and IdsE mediate strain-specific binding. Variants of FLAG-IdsE and IdsD-His$_6$ were subjected to pulldown assays and analyzed as described in the legend to Fig. 1. The IdsD and IdsE schematics are shaded according to the source strain, with the lighter shade representing variable regions. Interactions were tested as indicated in each panel. (A) IdsE variant FLAG-IdsE$_{BB\ to\ HI}$, in which the variable region of strain BB2000 was changed to that of strain HI4320, and IdsD-His$_6$ originating from either strain BB2000 or HI4320. (B) IdsD variant IdsD$_{small\ BB\ to\ HI}$-His$_6$, in which a small region of distinctive amino acids from strain BB2000 was changed to that of strain HI4320, and FLAG-IdsE originating from either strain BB2000 or HI4320, as well as the exchanged FLAG-IdsE$_{BB\ to\ HI}$. (C) IdsD variant IdsD$_{BB\ to\ HI}$-His$_6$, in which the extended variable region from strain BB2000 was changed to that of strain HI4320, and FLAG-IdsE originating from either strain BB2000 or HI4320, as well as the exchanged FLAG-IdsE$_{BB\ to\ HI}$.

the membrane-spanning region containing distinctive amino acids in IdsD from strain BB2000 (amino acids 777 to 865, where 16 of the 33 polymorphisms reside) with the analogous sequence from strain HI4320, resulting in IdsD$_{small\ BB\ to\ HI}$-His$_6$, expressed this protein in *E. coli*, and assayed for binding interactions by anti-FLAG immunoprecipitation. Surprisingly, we did not detect IdsD$_{small\ BB\ to\ HI}$-His$_6$ in pulldowns with FLAG-IdsE$_{BB}$, FLAG-IdsE$_{HI}$, or the VR-swap variant, FLAG-IdsE$_{BB\ to\ HI}$ (Fig. 3B). These data indicate that the exchanged residues in IdsD were sufficient to disrupt binding interactions but not sufficient to confer a different binding specificity.

Several additional amino acid polymorphisms flank the predicted IdsD variable region (Fig. 4). Using this as a basis, we introduced 2 amino acid exchanges (A761V and A765T) into the IdsD$_{small\ BB\ to\ HI}$-His$_6$ variant; this new construct, IdsD$_{BB\ to\ HI}$-His$_6$, was expressed in *E. coli* and assayed for binding interactions by anti-FLAG immunoprecipitation. IdsD$_{BB\ to\ HI}$-His$_6$ was detected in FLAG-IdsE$_{HI}$ but not in FLAG-IdsE$_{BB}$ pulldowns (Fig. 3C), indicating that conversion of these A761 and A765 residues in addition to the membrane-spanning variable region was sufficient to convert the binding specificity of IdsD from the IdsE variant of BB2000 to the IdsE variant of HI4320.

Strikingly, the IdsD$_{BB\ to\ HI}$-His$_6$ protein was also detected in FLAG-IdsE$_{BB\ to\ HI}$ protein pulldowns, indicating that exchanging both variable regions in tandem can rescue binding between the two hybrid proteins (Fig. 3C). Therefore, exchanging the variable region in either IdsD or IdsE from strain BB2000 to that of strain HI4320 is sufficient to abrogate interactions with a cognate binding partner and, importantly, to switch binding specificity to that of a foreign variant.

***In vitro* binding correlates with self-identity *in vivo*.** To determine whether the *in vitro* binding interactions correlated with *in vivo* behaviors, we used an *in vivo ids* expression system in which all *ids* genes are expressed from a plasmid under the native control of the *ids* promoter (pIds$_{BB}$) in a BB2000 mutant strain lacking the *ids* genes (Δ*ids*) (11). We chose this simplified system in which all other genes are identical except for the expressed *ids* genes so as to remove contributions to self-recognition-dependent boundary formation due to differences at other loci. To test the hypothesis that *in vitro* binding interactions correlate to *in vivo* self-identity, we replaced the now-defined variable regions (residues 761 to 865 of IdsD and 147 to 169 of IdsE) in the *idsD* and *idsE* genes, individually or together, in plasmid pIds$_{BB}$ with those from strain HI4320 and introduced each construct into the Δ*ids* strain.

These strains were subjected to boundary formation assays, which are currently the standard assay for studying self-identity in *P. mirabilis* (7–9, 11, 12). When two migrating populations merge to form a single swarm upon meeting, they are described as "self," and when a boundary forms between the two populations they are described as "nonself" (7, 11, 12). Expression of the *ids* genes from BB2000 (pIds$_{BB}$) in a Δ*ids* background results in a strain that merges with BB2000, indicating that BB2000 is recognized as self (11). In contrast, expression of the *ids* genes from strain HI4320 (pIds$_{HI}$) in a Δ*ids* background led to a boundary with the BB2000, HI4320, and Δ*ids* strains carrying pIds$_{BB}$ (Fig. 5; see Fig. S7 in the supplemental material). The boundary formation with strain HI4320 likely results from multiple factors independent from the *ids* genes, such as the putative cytotoxic *idr* (13) and/or *pef* (8) genes.

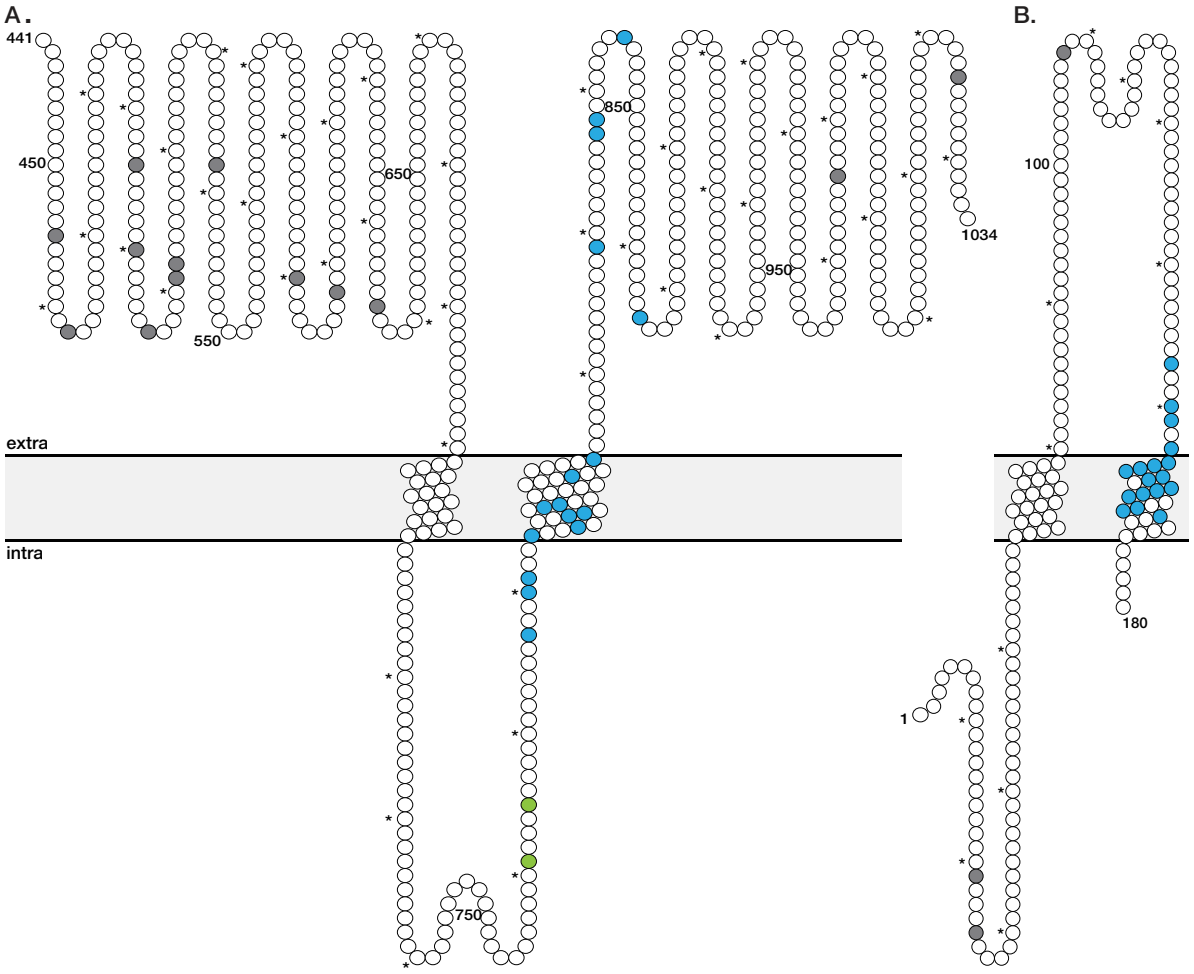

FIG 4 The IdsD and IdsE residues that mediate binding specificity overlap with predicted transmembrane domains. Presented are 2D projections (39) of $IdsD_{BB}$ from amino acids 441 to 1034 (A) and $IdsE_{BB}$ from amino acids 1 to 180 (B). Residues colored blue indicate BB2000 residues that were substituted for those from HI4320 in the $IdsD_{small\ BB\ to\ HI}$-His_6 and FLAG-$IdsE_{BB\ to\ HI}$ variants, while the residues colored in gray indicate amino acid polymorphisms between $IdsD_{BB}$ and $IdsD_{HI}$, as well as between $IdsE_{BB}$ and $IdsE_{HI}$, that were not exchanged. Positions A761 and A765 in IdsD ($IdsD_{BB\ to\ HI}$-His_6) are colored green. The membrane is colored light gray bordered by black lines.

Surprisingly, strains expressing an individual exchange, whether in *idsD* or in *idsE*, did not clearly merge or form a boundary with the strains expressing *ids* genes from either BB2000 or HI4320 and exhibited reduced swarm expansion (Fig. 5; see Fig. S7 in the supplemental material). Strikingly, the strain carrying *ids* alleles in which both the *idsD* and *idsE* variable regions were exchanged with those from strain HI4320, $pIds_{BB\text{-}idsD\text{-}BB\ to\ HI\text{-}idsE\text{-}BB\ to\ HI}$, merged with the strain expressing the *ids* genes from HI4320 ($pIds_{HI}$) and formed a boundary against the strain carrying the *ids* genes from BB2000 ($pIds_{BB}$) while also exhibiting a wild-type swarm expansion (Fig. 5; see Fig. S7 in the supplemental material). This observation is consistent with the observed *in vitro* binding interactions between the hybrid IdsD and IdsE proteins and the $IdsE_{HI}$ and $IdsD_{HI}$ proteins, respectively. From this, we conclude that the presence of cognate variable regions in both IdsD and IdsE, in otherwise isogenic strains, led to the conversion of strain-specific identity from that of one isolate to another *in vivo*, indicating that these binding interactions are one factor that contributes to the definition of strain identity.

DISCUSSION

Here we have shown that IdsD and IdsE bind to one another *in vitro* without the necessity of additional Ids or *P. mirabilis*-derived proteins (Fig. 1). Furthermore, we demonstrated that the *in vitro* IdsD-IdsE binding interaction is restricted by both allele (between proteins encoded within strain BB2000) and strain (between proteins encoded by strains BB2000, HI4320, and CW977), indicating the presence of allele-specific (i.e., cognate) IdsD and IdsE pairs in nature (Fig. 1 and 2). The information for the binding specificity between the IdsD and IdsE proteins is encoded in a short stretch of distinctive amino acids within each protein that comprises the variable region (Fig. 3 and 4). Strikingly, a positive binding interaction between IdsD and IdsE *in vitro* directly correlates with self-identity *in vivo* (Fig. 5).

The molecular recognition site between IdsD and IdsE overlaps with at least one predicted transmembrane domain for each protein (Fig. 4), suggesting that IdsD and IdsE may interact via an interface within the membrane. Exchange of residues within these

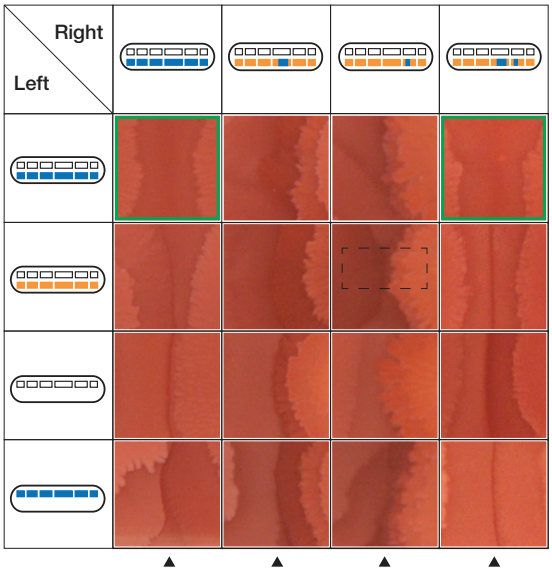

FIG 5 *In vitro* binding interactions between IdsD and IdsE correlate with self-identity *in vivo*. Boundary formation between strains expressing different IdsD-IdsE pairs was tested. Close-ups of contact regions between approaching swarms are shown. The arrowheads below indicate where two swarms meet, and the green outlines indicate swarms that have merged. The dashed box defines a region of contact between indicated swarms when more than two swarms are visible within the frame. Full images of swarm plates are shown in Fig. S7 in the supplemental material. Notations are as follows: ovals represent the tested swarm with the chromosomal *ids* locus at the top and the plasmid-contained *ids* locus at the bottom, and rectangles show six *ids* genes in sequence with alleles from strain BB2000 (orange), strain HI4320 (blue), or missing (white). Blue boxes within an orange box indicate a variable region exchange. The absence of bottom rectangles signifies that the strain carries the empty parent vector, pKG101.

transmembrane domains as well as in the predicted periplasmic loop was sufficient to disrupt native binding interactions for both IdsD and IdsE and was also sufficient to confer a new binding specificity for IdsE. Interestingly, for IdsD two additional predicted cytoplasmic residues (A761 and A765) were needed to convert binding specificity (Fig. 3 and 4), suggesting that the variable region for IdsD might be extended to the regions flanking the transmembrane domain. These two residues may also contribute to the stability of IdsD or to the fold of the variable region.

Binding specificity between two proteins is crucial for many intracellular processes, e.g., bacterial histidine kinase (HK) and response regulator (RR) proteins (reviewed in reference 26). Multiple variants of homologous HK and RR proteins are simultaneously present in a single *Caulobacter crescentus* cell, yet signaling via phosphorylation is restricted to cognate HK-RR protein pairs that are often encoded by adjacent genes (27). Specific residues in the HK and RR proteins define the specificity of these interactions, restricting the ability of a given protein to interact with a noncognate variant and permitting the predictive redesign of a protein's signaling specificity (28–33). As the specificity of signaling interactions is sufficient to alter intracellular processes, it stands to reason that variant-specific binding interactions between proteins may also drive population and social behaviors. Since no additional proteins are required for the IdsD-IdsE interaction, we posit that the binding between these two proteins is the central protein-protein interaction determining *P. mirabilis* Ids-mediated self-recognition and most likely occurs in the cell envelope.

In fact, the *in vitro* binding specificity between IdsD and IdsE appears to be preferential to not only an endogenous *idsE* variant but also the *idsE* allele immediately adjacent to *idsD* (see Table S1 in the supplemental material). The role of the IdsE-like proteins encoded by nonadjacent alleles remains unknown. These orphan IdsE variants may interact with foreign IdsDs from the environment or may serve as reservoirs for alternate identity by replacing the canonical *idsE* gene via allelic exchange. Interestingly, not all *P. mirabilis* genomes contain *ids* genes, and it is unclear whether this is due to a recent acquisition of *ids* genes within the *P. mirabilis* species or whether the *ids* genes have been lost in some strains. Alteri et al. have demonstrated that a HI4320-derived strain with a disruption in the *idsD* gene is able to merge with its parental strain; however, the effect of a full *ids* deletion or a deletion of *idsE* in strain HI4320 remains to be examined (8). The diversity of IdsD and IdsE likely extends beyond *P. mirabilis*, as genes with similarity to *idsD* and *idsE* are adjacent to each other in other bacterial species (see Fig. S8 in the supplemental material [22–25]), raising the possibilities that IdsD and IdsE may encode strain-specific information in other species and that IdsD and IdsE may coevolve. The specific binding between adjacently encoded IdsD and IdsE pairs observed here supports a hypothesis for selective pressure on *idsD* and *idsE* to maintain a complementary protein interaction interface. It remains to be determined how each distinctive amino acid contributes to binding specificity and how the remaining portions of the IdsD and IdsE proteins contribute to biological function.

Here we have reported new biochemical information on two self-identity proteins of unknown function and structure. This information is a necessary contribution for developing a mechanistic model of self-recognition, as well as for a fuller understanding of protein-protein interactions in *P. mirabilis* and other bacteria. However, many questions remain unresolved. For example, multiple modes for the IdsD and IdsE interaction *in vivo* are possible. The IdsD and IdsE interactions could occur (i) between neighboring cells, (ii) within a single cell, or (iii) through a combination of both.

We prefer a model in which IdsD from one cell is communicated to a neighboring self cell; a positive binding interaction between the transferred IdsD and the recipient's encoded cognate IdsE would then cause a signaling cascade in the recipient cell, ultimately resulting in behaviors that are beneficial for kin, such as swarming, which is a cooperative method of motility. In support of this hypothesis, we have observed reduced motility in strains in which only the IdsD and IdsE proteins are noncognate and thus do not bind *in vitro* (see Fig. S7 in the supplemental material). Given that the T6S system is necessary for boundary formation (8, 13) and for the export of IdsD from liquid-grown cells into the extracellular medium (13), IdsD may be transported directly into (or onto) a neighboring cell via the T6S system to elicit a response. Indeed, T6S systems in *Pseudomonas aeruginosa* can transfer macromolecules known as effectors directly from one cell into its neighbor, resulting in the recipient's cell death (16). However, we have not observed transfer of IdsD directly into neighboring cells in liquid or on surfaces and as such cannot definitively conclude whether IdsD is transported into neighboring cells. Further, the presence of IdsD in the extracellular supernatant despite the predicted transmembrane domains (13) raises the question of whether multiple isoforms of IdsD are present within a cell: a cell envelope-localized isoform and an exported isoform. Alternatively, since IdsD and IdsE

each have two predicted transmembrane domains, IdsD and IdsE could potentially form an envelope-spanning complex within a single cell that in turn interacts with a similar complex in a neighboring cell. Nonetheless, it remains likely that adjacent cells share identity information through the actions of IdsD and IdsE. Future work will be aimed at determining the topology, subcellular localization, and three-dimensional structure of the native IdsD-IdsE complex.

Without knowledge of the topology for IdsD-IdsE binding interactions *in vivo*, we cannot predict how these interactions contribute to boundary formation. However, these data are consistent with our current model for self-recognition in *P. mirabilis* strain BB2000, in which the absence of cognate Ids proteins denotes that the interacting cell is missing self-identifiers (13). Defining self-identity is at the foundation of many group behaviors mediated by self- versus nonself-recognition. Features of *P. mirabilis* self-recognition are shared with recognition systems like those of other social microbes, e.g., kin-specific binding interactions between proteins to define identity. In this study, we definitively link *in vitro* binding affinity with an *in vivo* self-recognition behavior. We have also begun to map the recognition interface necessary for the allele-specific binding between two self-identity proteins. Better understanding of molecular recognition among proteins, such as this one that drives population identity, may provide insights into how genetic mutations and genomic evolution emerge into or are constrained by population behaviors.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. All strains and plasmid constructions are described in the supplemental material (see Table S2). *E. coli* and *P. mirabilis* strains were maintained on LB and LSW agar, respectively (34). CM55 blood agar base agar (Oxoid, Basingstoke, England) was used for swarm colony growth. All strains were grown in LB broth under aerobic conditions at 16, 30, or 37°C. Antibiotics were used at the following concentrations: carbenicillin, 100 μg/ml; tetracycline, 15 μg/ml; kanamycin, 35 μg/ml; and chloramphenicol, 50 μg/ml.

Boundary assays. All boundary assays were performed as previously described (13) on swarm-permissive nutrient plates with kanamycin.

FLAG immunoprecipitations from *P. mirabilis* cell extracts. *P. mirabilis* strains carrying pIds plasmids were inoculated from overnight cultures onto three swarm agar plates and incubated for ~20 h until the population almost reached the edge of the petri dish. Cells were resuspended in LB, harvested by centrifugation, and stored at −80°C. Pellets were resuspended in 1 ml cell lysis buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1% Triton X-100, 1 mM EDTA) supplemented with Complete protease inhibitor cocktail (Roche, Basel, Switzerland) and lysed by vortexing with cell disruptor beads (0.1-mm diameter; Electron Microscopy Sciences, Hatfield, PA). Lysates were cleared by centrifugation and applied to 20 μl preequilibrated anti-FLAG M2 antibody resin (Sigma-Aldrich, St. Louis, MO). Control lysate (not containing a FLAG-tagged protein) was supplemented with 2 μg of FLAG-BAP protein (Sigma-Aldrich, St. Louis, MO). Lysates were incubated with resin for 2 h at 4°C. Unbound cell extract was removed. Resin was washed five times in wash buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1% Triton X-100), and bound proteins were eluted with 50 μl of elution buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 200 ng/μl 3× FLAG peptide) for 45 min at 4°C. The elution was recentrifuged, and the top 40 μl (of the original 50 μl) was retained. Samples of load (L), flowthrough (F) (i.e., proteins in supernatant after incubation with resin), and elution (E) were separated by SDS-PAGE and analyzed by Western blotting.

Co-immunoprecipitation assays from *E. coli* cell extracts. BL21(DE3) cells (Invitrogen Corporation, Carlsbad, CA) transformed with overexpression plasmids were grown in 25 ml of LB with 100 μg/ml carbenicillin under shaking conditions at 30°C for 3 h, cooled on ice, induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside), and incubated overnight while shaking at 16°C. Cells were harvested by centrifugation and then stored at −80°C. Cells were lysed as described above, and the total protein concentration was assessed by a microplate reader-based Bradford assay (Bio-Rad Laboratories, Hercules, CA). Seven hundred microliters of FLAG-tagged extracts (normalized with lysis buffer to equivalent protein concentrations) was mixed with 300 μl of the His_6-tagged extracts, applied to anti-FLAG resin, immunoprecipitated, and analyzed as described above.

SDS-PAGE and Western blots. Samples from the immunoprecipitation assays described above were separated by gel electrophoresis using 12% or 15% Tris-Tricine polyacrylamide gels, transferred to nitrocellulose membranes, and probed with rabbit anti-IdsD (1:2,000), rabbit anti-IdsE (1:2,000), rabbit anti-FLAG (1:4,000; Sigma-Aldrich, St. Louis, MO), rabbit anti-His_6 (1:2,000; Abcam, Cambridge, England), or mouse anti-sigma-70 (1:1,000; Thermo, Fisher Scientific, Waltham, MA), followed by goat anti-rabbit conjugated to horseradish peroxidase (HRP) (1:5,000; KPL, Inc., Gaithersburg, MD) or goat anti-mouse conjugated to HRP (1:5,000; KPL, Inc., Gaithersburg, MD) and developed with Immun-Star HRP substrate kit (Bio-Rad Laboratories, Hercules, CA). Antibodies specific to IdsD amino acids 4 to 18 (EVNEKYLTPQERKAR) and IdsE amino acids 298 to 312 (EQILAKLDQEKEHHA) were raised in rabbits using standard protocols (Covance, Dedham, MA). Blots were visualized using a Chemidoc (Bio-Rad Laboratories, Hercules, CA). JPEG images of blots were converted to TIFF files using Adobe Photoshop (Adobe Systems, San Jose, CA), and figures were made in Adobe Illustrator (Adobe Systems, San Jose, CA).

Bioinformatics analysis and construction of 2D projection graphics. Bioinformatics analysis of the IdsD and IdsE amino acid sequences from strains BB2000 and HI4320 were performed using the web interfaces of PredictProtein (35), TMFinder (36), Hmmer (37), and Phyre^2 (38). Two-dimensional (2D) projections of $IdsD_{BB}$ and $IdsE_{BB}$ were prepared using the web-accessible Protter software (http://wlab.ethz.ch/protter/start/) (39), which employs Phobius (40, 41) to predict transmembrane domains and orientation. Colors and red lines were added using Adobe Illustrator (Adobe Systems, San Jose, CA). The sequence alignment methods are described in the supplemental material.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00251-15/-/DCSupplemental.

Figure S1, EPS file, 2.6 MB.
Figure S2, EPS file, 0.9 MB.
Figure S3, TIF file, 2.4 MB.
Figure S4, TIF file, 2.5 MB.
Figure S5, EPS file, 0.4 MB.
Figure S6, EPS file, 1.1 MB.
Figure S7, TIF file, 2.7 MB.
Figure S8, EPS file, 0.7 MB.
Table S1, DOCX file, 0.1 MB.
Table S2, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS

We thank members of the Gibbs laboratory for experimental tools and advice, as well as Richard Losick, Bodo Stern, and Alexander Schier for comments on the manuscript.

The Canadian Institutes for Health Research (L.C.), Harvard University, the George W. Merck Fund, and the David and Lucile Packard Foundation, funded our research.

REFERENCES

1. **Hirose S, Benabentos R, Ho HI, Kuspa A, Shaulsky G.** 2011. Self-recognition in social amoebae is mediated by allelic pairs of tiger genes. Science **333:**467–470. http://dx.doi.org/10.1126/science.1203903.
2. **Benabentos R, Hirose S, Sucgang R, Curk T, Katoh M, Ostrowski EA, Strassmann JE, Queller DC, Zupan B, Shaulsky G, Kuspa A.** 2009.

Polymorphic members of the lag gene family mediate kin discrimination in *Dictyostelium*. Curr Biol **19:**567–572. http://dx.doi.org/10.1016/j.cub.2009.02.037.
3. **Ducret A, Fleuchot B, Bergam P, Mignot T.** 2013. Direct live imaging of cell-cell protein transfer by transient outer membrane fusion in *Myxococcus xanthus*. Elife **2:**e00868. http://dx.doi.org/10.7554/eLife.00868.
4. **Nudleman E, Wall D, Kaiser D.** 2005. Cell-to-cell transfer of bacterial outer membrane lipoproteins. Science **309:**125–127. http://dx.doi.org/10.1126/science.1112440.
5. **Pathak DT, Wei X, Bucuvalas A, Haft DH, Gerloff DL, Wall D.** 2012. Cell contact-dependent outer membrane exchange in myxobacteria: genetic determinants and mechanism. PLoS Genet **8:**e1002626. http://dx.doi.org/10.1371/journal.pgen.1002626.
6. **Pathak DT, Wei X, Dey A, Wall D.** 2013. Molecular recognition by a polymorphic cell surface receptor governs cooperative behaviors in bacteria. PLoS Genet **9:**e1003891. http://dx.doi.org/10.1371/journal.pgen.1003891.
7. **Dienes L.** 1947. Further observations on the reproduction of bacilli from large bodies in *Proteus* cultures. Proc Soc Exp Biol Med **66:**97. http://dx.doi.org/10.3181/00379727-66-15994.
8. **Alteri CJ, Himpsl SD, Pickens SR, Lindner JR, Zora JS, Miller JE, Arno PD, Straight SW, Mobley HL.** 2013. Multicellular bacteria deploy the type VI secretion system to preemptively strike neighboring cells. PLoS Pathog **9:**e1003608. http://dx.doi.org/10.1371/journal.ppat.1003608.
9. **Budding AE, Ingham CJ, Bitter W, Vandenbroucke-Grauls CM, Schneeberger PM.** 2009. The Dienes phenomenon: competition and territoriality in swarming *Proteus mirabilis*. J Bacteriol **191:**3892–3900. http://dx.doi.org/10.1128/JB.00975-08.
10. **Gibbs KA, Greenberg EP.** 2011. Territoriality in *Proteus*: advertisement and aggression. Chem Rev **111:**188–194. http://dx.doi.org/10.1021/cr100051v.
11. **Gibbs KA, Urbanowski ML, Greenberg EP.** 2008. Genetic determinants of self identity and social recognition in bacteria. Science **321:**256–259. http://dx.doi.org/10.1126/science.1160033.
12. **Senior BW.** 1977. The Dienes phenomenon: identification of the determinants of compatibility. J Gen Microbiol **102:**235–244. http://dx.doi.org/10.1099/00221287-102-2-235.
13. **Wenren LM, Sullivan NL, Cardarelli L, Septer AN, Gibbs KA.** 2013. Two independent pathways for self-recognition in *Proteus mirabilis* are linked by type VI-dependent export. mBio **4**(4):e00374-13. http://dx.doi.org/10.1128/mBio.00374-13.
14. **Russell AB, Peterson SB, Mougous JD.** 2014. Type VI secretion system effectors: poisons with a purpose. Nat Rev Microbiol **12:**137–148. http://dx.doi.org/10.1038/nrmicro3185.
15. **Koskiniemi S, Lamoureux JG, Nikolakakis KC, t'Kint de Roodenbeke C, Kaplan MD, Low DA, Hayes CS.** 2013. Rhs proteins from diverse bacteria mediate intercellular competition. Proc Natl Acad Sci U S A **110:**7032–7037. http://dx.doi.org/10.1073/pnas.1300627110.
16. **Russell AB, Hood RD, Bui NK, LeRoux M, Vollmer W, Mougous JD.** 2011. Type VI secretion delivers bacteriolytic effectors to target cells. Nature **475:**343–347. http://dx.doi.org/10.1038/nature10244.
17. **Miyata ST, Kitaoka M, Brooks TM, McAuley SB, Pukatzki S.** 2011. *Vibrio cholerae* requires the type VI secretion system virulence factor VasX to kill *Dictyostelium discoideum*. Infect Immun **79:**2941–2949. http://dx.doi.org/10.1128/IAI.01266-10.
18. **Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ.** 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. Proc Natl Acad Sci U S A **104:**15508–15513. http://dx.doi.org/10.1073/pnas.0706532104.
19. **Salomon D, Kinch LN, Trudgian DC, Guo X, Klimko JA, Grishin NV, Mirzaei H, Orth K.** 2014. Marker for type VI secretion system effectors. Proc Natl Acad Sci U S A **111:**9271–9276. http://dx.doi.org/10.1073/pnas.1406110111.
20. **Gibbs KA, Wenren LM, Greenberg EP.** 2011. Identity gene expression in *Proteus mirabilis*. J Bacteriol **193:**3286–3292. http://dx.doi.org/10.1128/JB.01167-10.
21. **Sullivan NL, Septer AN, Fields AT, Wenren LM, Gibbs KA.** 2013. The complete genome sequence of *Proteus mirabilis* strain BB2000 reveals differences from the *P. mirabilis* reference strain. Genome Announc **1:**e00024-13. http://dx.doi.org/10.1128/genomeA.00024-13.
22. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947–2948. http://dx.doi.org/10.1093/bioinformatics/btm404.
23. **Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R.** 2010. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res **38:**W695–W699. http://dx.doi.org/10.1093/nar/gkq313.
24. **McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R.** 2013. Analysis tool Web services from the EMBL-EBI. Nucleic Acids Res **41:**W597–W600. http://dx.doi.org/10.1093/nar/gkt376.
25. **Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ.** 2009. Jalview, version 2. a multiple sequence alignment editor and analysis workbench. Bioinformatics **25:**1189–1191. http://dx.doi.org/10.1093/bioinformatics/btp033.
26. **Stock AM, Robinson VL, Goudreau PN.** 2000. Two-component signal transduction. Annu Rev Biochem **69:**183–215. http://dx.doi.org/10.1146/annurev.biochem.69.1.183.
27. **Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT.** 2005. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. PLoS Biol **3:**e334. http://dx.doi.org/10.1371/journal.pbio.0030334.
28. **Podgornaia AI, Casino P, Marina A, Laub MT.** 2013. Structural basis of a rationally rewired protein-protein interface critical to bacterial signaling. Structure **21:**1636–1647. http://dx.doi.org/10.1016/j.str.2013.07.005.
29. **Capra EJ, Perchuk BS, Lubin EA, Ashenberg O, Skerker JM, Laub MT.** 2010. Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. PLoS Genet **6:**e1001220. http://dx.doi.org/10.1371/journal.pgen.1001220.
30. **Casino P, Rubio V, Marina A.** 2009. Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. Cell **139:**325–336. http://dx.doi.org/10.1016/j.cell.2009.08.032.
31. **Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT.** 2008. Rewiring the specificity of two-component signal transduction systems. Cell **133:**1043–1054. http://dx.doi.org/10.1016/j.cell.2008.04.040.
32. **Weigt M, White RA, Szurmant H, Hoch JA, Hwa T.** 2009. Identification of direct residue contacts in protein-protein interaction by message passing. Proc Natl Acad Sci U S A **106:**67–72. http://dx.doi.org/10.1073/pnas.0805923106.
33. **Bell CH, Porter SL, Strawson A, Stuart DI, Armitage JP.** 2010. Using structural information to change the phosphotransfer specificity of a two-component chemotaxis signalling complex. PLoS Biol **8:**e1000306. http://dx.doi.org/10.1371/journal.pbio.1000306.
34. **Belas R, Erskine D, Flaherty D.** 1991. Transposon mutagenesis in *Proteus mirabilis*. J Bacteriol **173:**6289–6293.
35. **Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigschmid P, Schafferhans A, Roos M, Bernhofer M, Richter L, Ashkenazy H, Punta M, Schlessinger A, Bromberg Y, Schneider R, Vriend G, Sander C, Ben-Tal N, Rost B.** 2014. PredictProtein—an open resource for online prediction of protein structural and functional features. Nucleic Acids Res **42:**W337–W343. http://dx.doi.org/10.1093/nar/gku366.
36. **Deber CM, Wang C, Liu LP, Prior AS, Agrawal S, Muskat BL, Cuticchia AJ.** 2001. TM finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. Protein Sci **10:**212–219. http://dx.doi.org/10.1110/ps.30301.
37. **Finn RD, Clements J, Eddy SR.** 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res **39:**W29–W37. http://dx.doi.org/10.1093/nar/gkr367.
38. **Kelley LA, Sternberg MJ.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc **4:**363–371. http://dx.doi.org/10.1038/nprot.2009.2.
39. **Omasits U, Ahrens CH, Müller S, Wollscheid B.** 2014. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics **30:**884–886. http://dx.doi.org/10.1093/bioinformatics/btt607.
40. **Käll L, Krogh A, Sonnhammer EL.** 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res **35:**W429–W432. http://dx.doi.org/10.1093/nar/gkm256.
41. **Käll L, Krogh A, Sonnhammer EL.** 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol **338:**1027–1036. http://dx.doi.org/10.1016/j.jmb.2004.03.016.